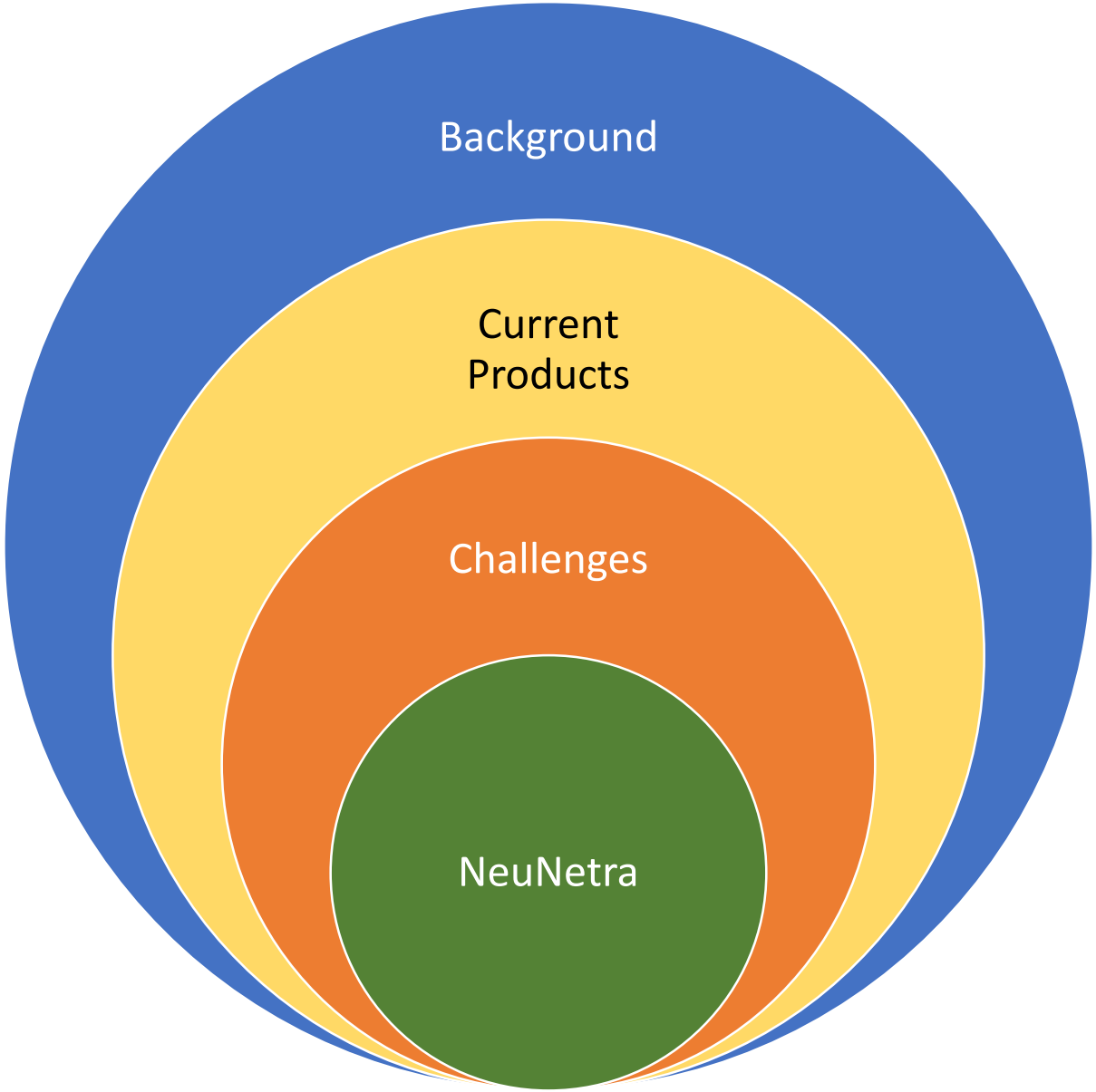
AI/ML at the edge

Prasad Panchangam

Founder CEO, NeuNetra

Dec 2019

Agenda



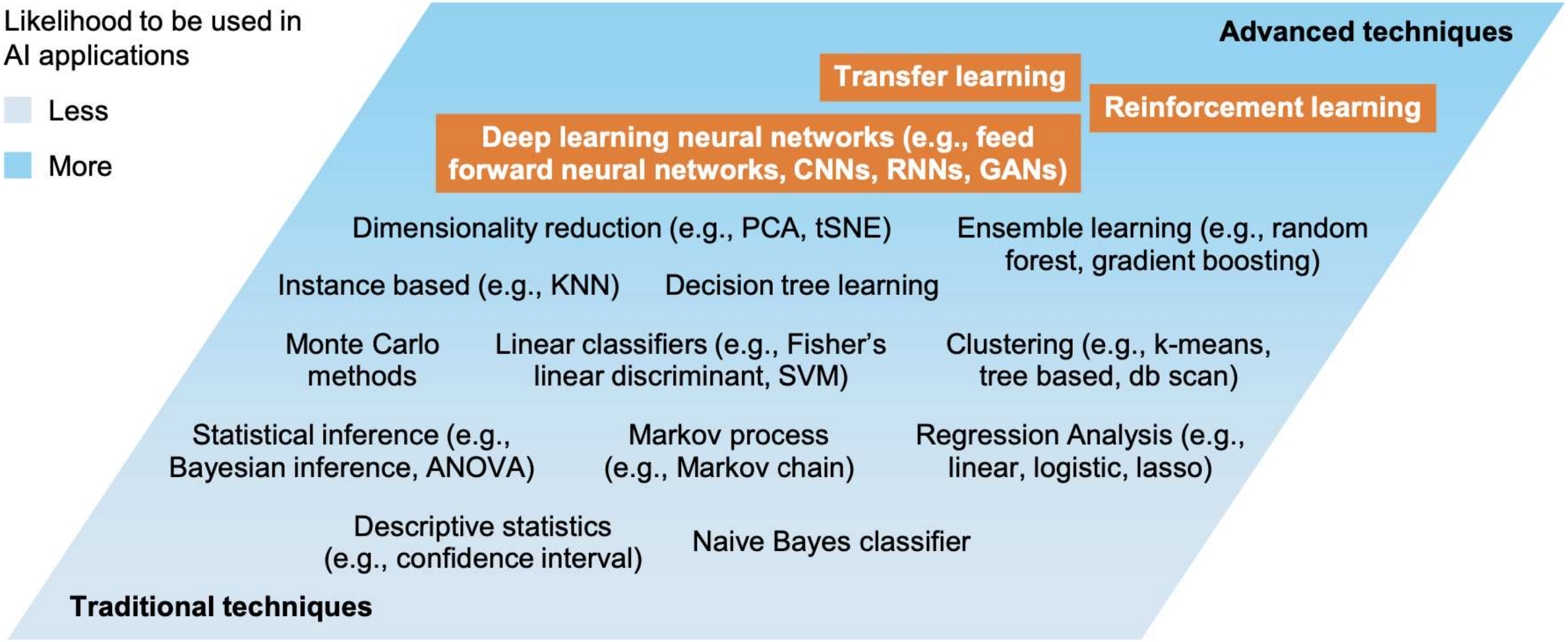
Background and Definitions

Intelligence moving to the edge

Key benefits of intelligence at the edge:

Low Latency	Low Power	Low Cost	High Privacy	High Reliability

AI, ML and other analytics techniques



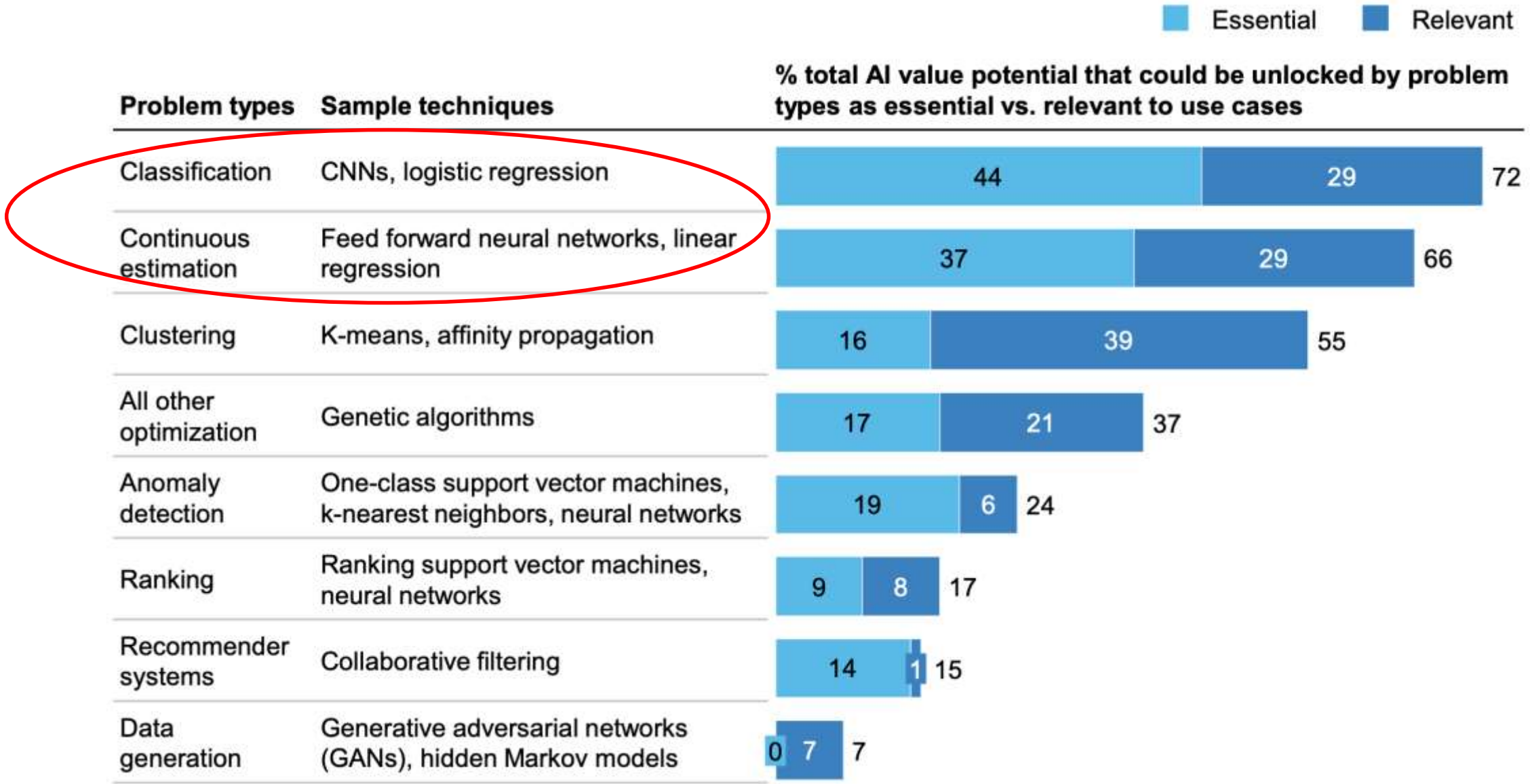
Definitions for this presentation (AI on the edge)

- AI - “deep learning” techniques using artificial neural networks - can be used to solve a variety of problems.

- TECHNIQUES – those that address classification and estimation problems - currently the most widely applicable for the edge

- FOCUS — feed forward neural networks, recurrent neural networks, and convolutional neural networks — Potentially enable the creation of between $3.5 trillion and $5.8 trillion in value annually. (says McKinsey)

Problem types and sample techniques



Current Products

Smartphones add AI engines

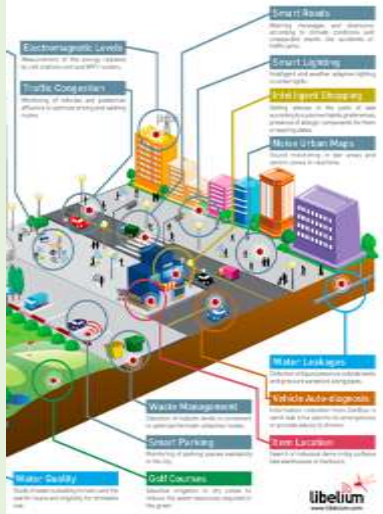
- Apple A11, A12 integrate neural engine
- Samsung galaxy S9 – neural engine from DeePhi
- Huawei Kirin 970, 980 – Neural engines from cambricon
- Qualcomm Snapdragon 845, 855 – Hexagon vector DSP
- Mediatek P90 – Cadence P6 plus custom neural engine
- Trickling down to mid-tier phones

- These engines enable “new applications” such as 3D face recognition
- More efficient than CPU or GPU

AI in IoT

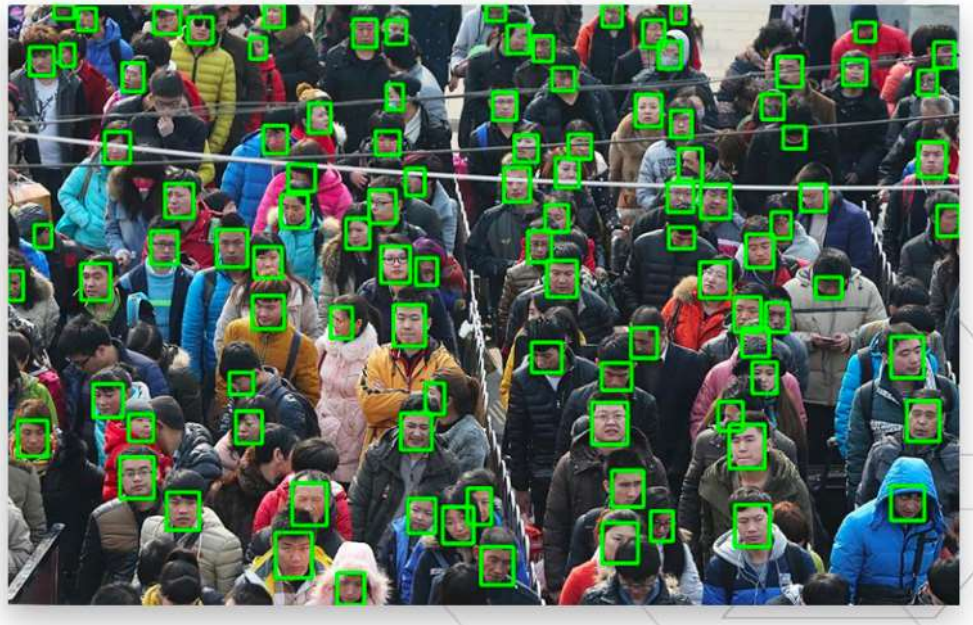
- Voice assistants
- Security cameras
- Drones

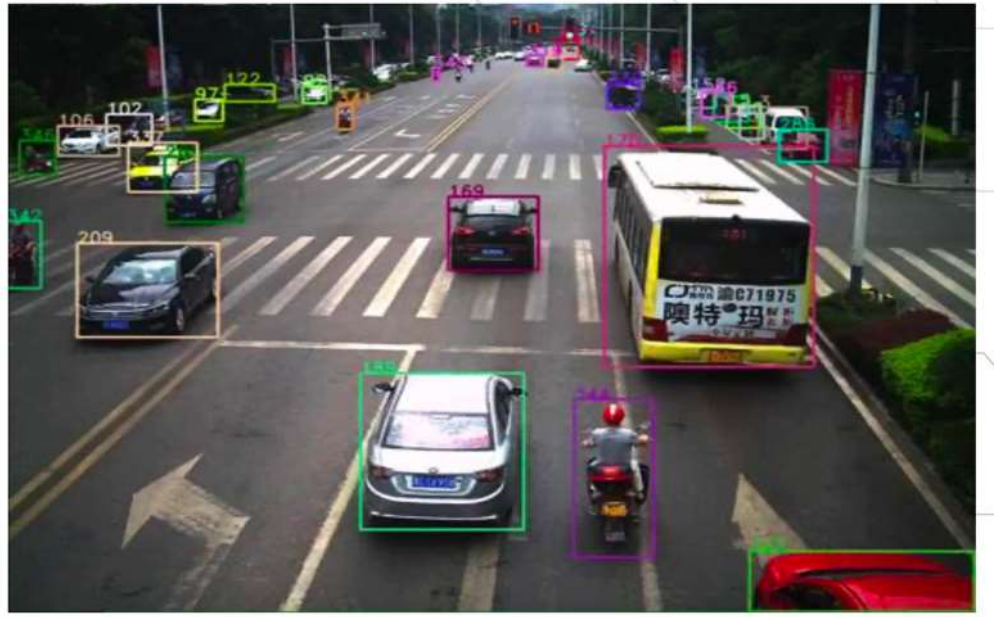
- Industrial IoT is booming
  - Smart buildings, hospitals,
  - factories, farms, infrastructure
- Consumer IoT is starting to take off
  - Wearables, Smart homes, video doorbells



Target – Smart cameras

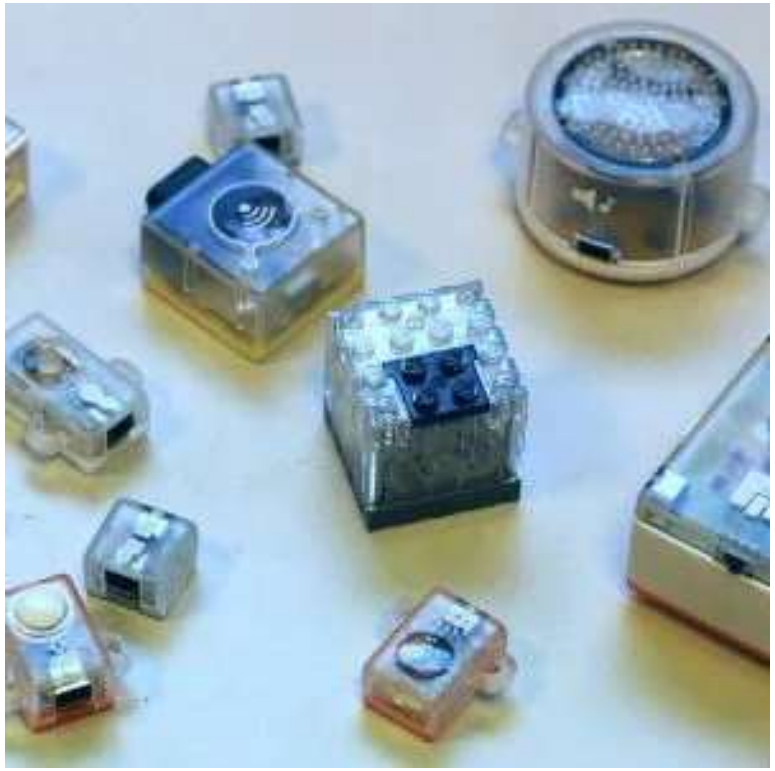
- Custom HW enables more complex DNNs than a CPU can handle
- Surveillance market is booming
  - Especially in China
- Intel Myriad X, Bitmain BM1880 offer 1TOPS at 2.5W
- Other Chinese chip makers:
  - Canaan/Kendryte, Cambricon, HiSilicon, Horizon robotics





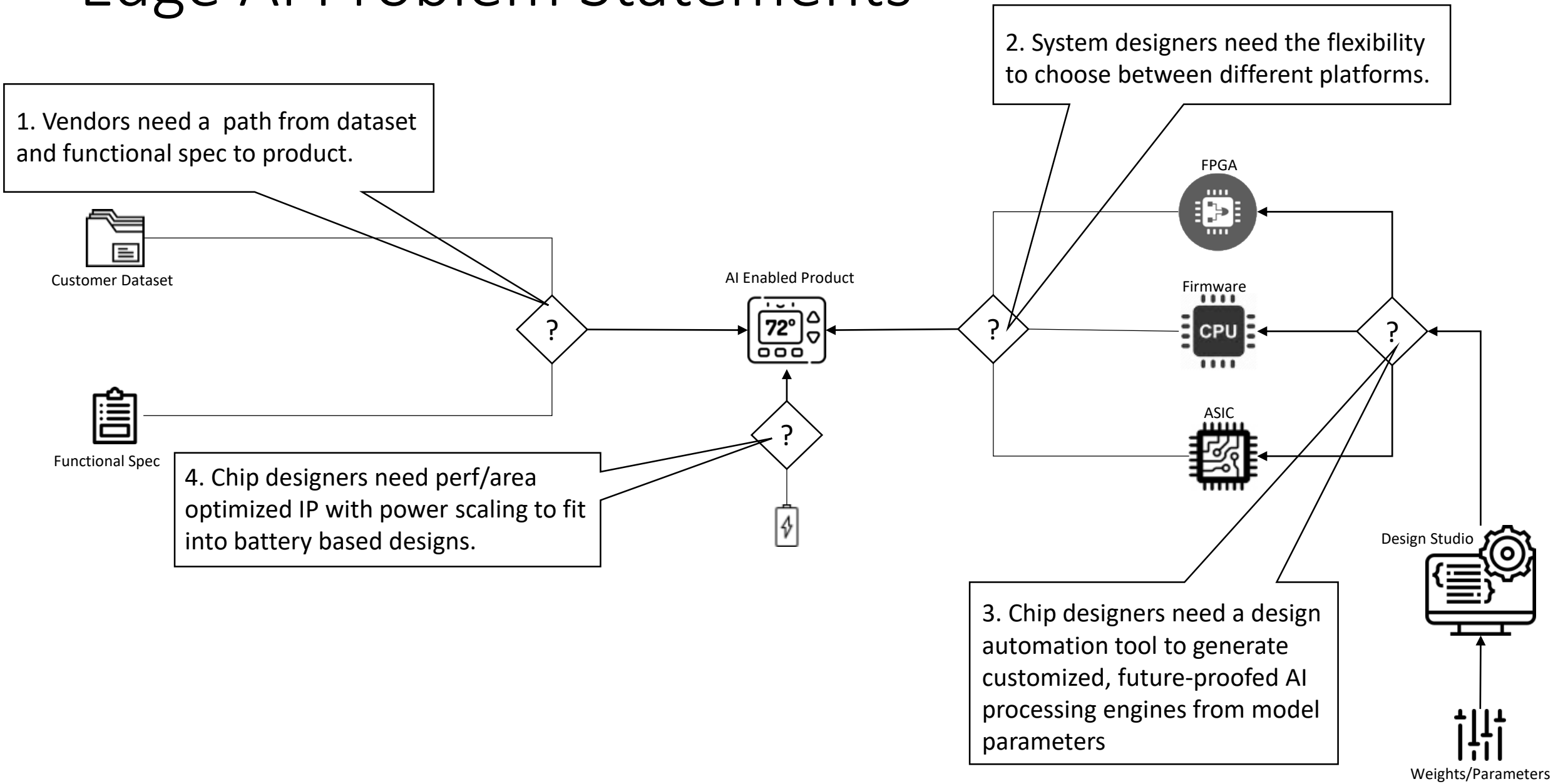
Tiny AI engines for IoT sensors

- Many IoT devices today use cloud processing
- Microcontroller CPUs can handle simple neural networks
  - Cortex M4 has DSP extensions to improve DNN perf
  - ARM CMSIS-NN includes tools for porting DNNs to its CPUs
- Small IoT devices benefit from AI accelerators
  - Extends battery life
- Eta compute offers Tensai MCU with coolflux DSP accelerator
  - Handles basic voice recognition at 2mW (15x less energy than standard CPU)
- Greenwaves GAP8 MCU implements 8 core accelerator using RISC-V
  - Scales from 4mW to 70mW, can handle voice/image recognition
- Syntiant
  - Analog NDP



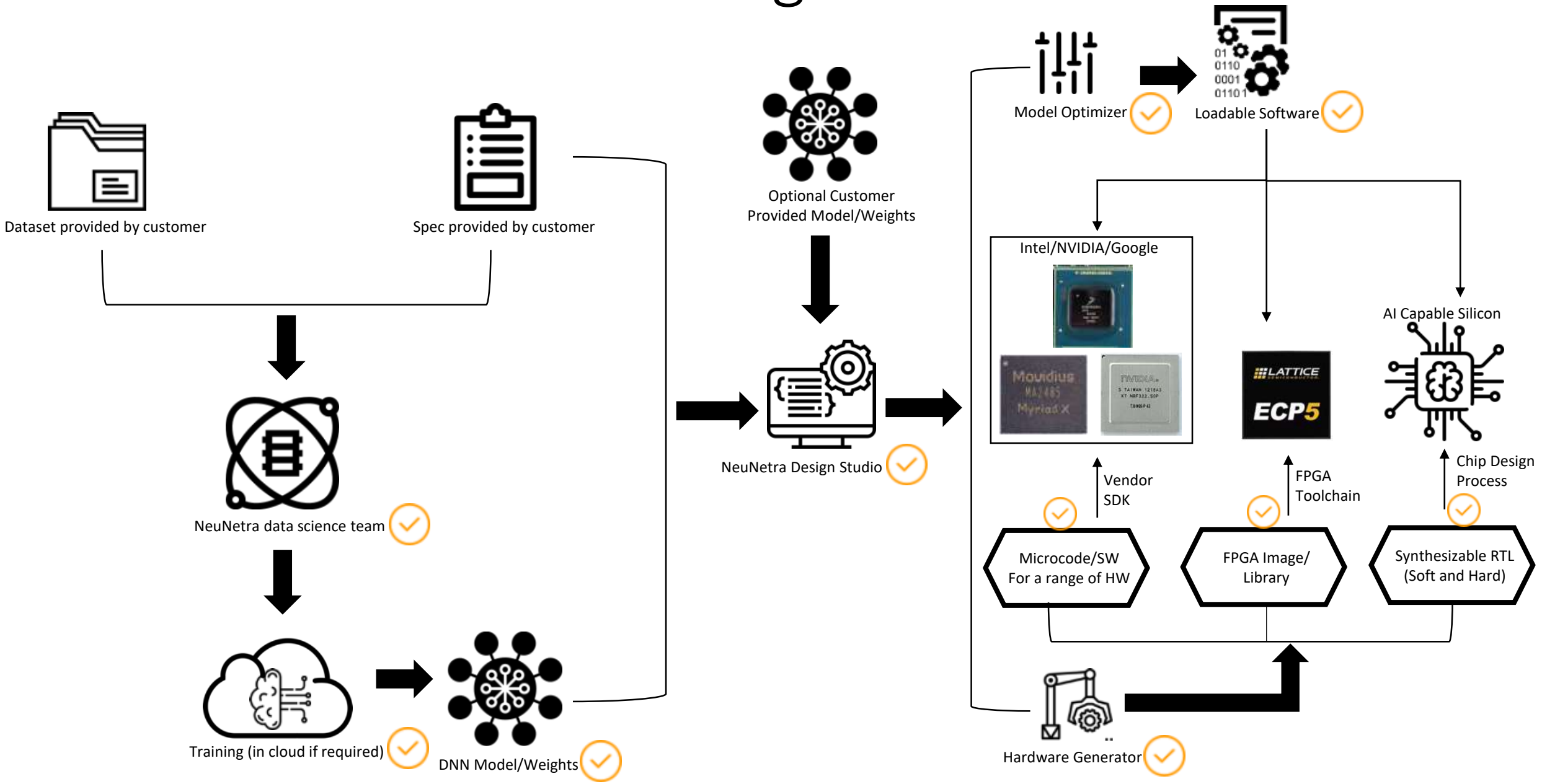
Challenge

Edge AI Problem Statements



NeuNetra

The NeuNetra advantage



NeuNetra Unique Value/Positioning

- ✓ End-to-End (data to product) capability
- ✓ Fully functional, Stand-alone, Customizable and Extensible AI processing engines
- ✓ Platform independent. Can generate for ASIC, FPGA, Intel, Google, NVIDIA, Flexlogix
- ✓ AI Engines come with critical optimizations and power scaling capabilities to support sensor level applications
- ✓ Design Studio based solution

Thank You!

prasad@cloudxim.com

+91-99000-42739